SISTEMAS DE DETERMINACIÓN Y EVALUACIÓN DE METAS DE LOGROS DE APRENDIZAJE ESCOLAR COMO INSTRUMENTOS PARA MEJORAR LA CALIDAD, LA EQUIDAD Y LA RESPONSABILIZACIÓN EN LOS PROCESOS EDUCATIVOS EN AMÉRICA LATINA[1]

Patricia Arregui[2]

Agosto del 2000

1. Introducción

En los últimos diez años, se ha venido ampliando y consolidando un consenso respecto a que es indispensable que los niños, jóvenes y adultos mejoren sus capacidades y logros de aprendizaje, tanto en la escuela, como en todos los otros espacios y momentos de sus vidas.

El conocimiento y la capacidad de aprender se consideran cada vez más centrales para la calidad de vida y el bienestar de los individuos y sus comunidades. Se perciben también como esenciales para la preservación de las civilizaciones y las culturas y para las posibilidades de una participación activa y beneficiosa de toda la diversidad de personas y poblaciones que habitan en un mundo y una economía globalizados, crecientemente dependientes del crecimiento y del cambio de los saberes. Y, a la vez que se reconoce la multiplicación y dispersión de los agentes que efectivamente "educan" a las nuevas generaciones, se refuerza el rol de la escuela como actor central y portador principal de la misión de promover el desarrollo de las capacidades cognitivas y otras esenciales para la vida y convivencia humanas.

Lo que dista aún bastante de poder considerarse materia de consenso es *cuáles son esas capacidades esenciales que la escuela puede y debe contribuir a generar, cómo se puede saber si los educandos las están efectivamente desarrollando, cuál es o debe ser la contribución específica que los distintos agentes educativos (familias, maestros, funcionarios, políticos, los mismos alumnos) deben hacer para lograrlos y cómo puede y debe estimularse, recompensarse o asignar otro tipo de consecuencias al cumplimiento o no de las respectivas responsabilidades*.

No obstante la diversidad de opiniones vigentes al respecto, a lo largo del tiempo, y particularmente en las décadas recientes, en muchos países con tradiciones diversas respecto a la estructura de sus sistemas educativos se ha generado *estándares educacionales* (con ese u otros nombres) que definen lo que los estudiantes deben saber y saber hacer, así como cuán bien deben hacerlo. Se espera que esos estándares operen como anclas o, mejor aun, como ejes articuladores e impulsores de otros componentes o acciones para el mejoramiento educativo permanente (el desarrollo curricular y de textos y otros materiales educativos, la

---

[1] Documento elaborado a solicitud de UNESCO/OREALC, para su Estudio Prospectivo sobre la Educación en América Latina y el Caribe y el seminario correspondiente a realizarse en Santiago de Chile en agosto del 2000.

[2] La autora es investigadora del Grupo de Análisis para el Desarrollo (GRADE) y coordinadora del Grupo de Trabajo sobre Estándares y Evaluación del Programa de Promoción de la Reforma Educativa en América Latina (GTEE-PREAL). Agradece los comentarios a un primer borrador del mismo ofrecidos por Guillermo Ferrer. Las ideas y opiniones contenidas en el documento son responsabilidad propia y no implican una posición institucional al respecto de GRADE ni del PREAL.

formación y capacitación en servicio de los docentes, el diseño y utilización de evaluaciones de logros, etc.). En muchos países desarrollados, existen mecanismos formales diversos para evaluar su logro en distintos momentos de la vida escolar de niños y jóvenes.

En la gran mayoría de los países de América Latina, en cambio, si bien han habido muchos y significativos avances en los últimos años, todavía no son "sentidos comunes" cuestiones tales como la conveniencia de mantener expectativas académicas altas y claramente formuladas para todos los escolares, medir de manera pertinente y rigurosa el logro de esas expectativas, utilizar los resultados de esas mediciones para mejorar las políticas, programas y acciones educativas, explicitar la asignación de responsabilidades a distintos grupos de agentes del proceso educativo y exigir evidencias del cumplimiento exitoso de esas responsabilidades. En términos más generales, podría decirse que aún no está del todo enraizada en la conciencia pública la necesidad de políticas que permitan el establecimiento de metas compartidas y la responsabilización también compartida por los resultados.

Las demandas y tendencias mencionadas han sido recogidas e impulsadas por una ya larga serie de documentos de posición de diversas organizaciones internacionales y regionales. Así, por ejemplo, el informe Delors[3] preparado a solicitud de la UNESCO, aunque no abundó en el tema, no dejó de resaltar la importancia de establecer un sistema de evaluación pública imparcial. Argumentaba que la magnitud de los recursos asignados al sector justifican que para comprometer inversiones adicionales en el mismo sea necesario establecer procedimientos de rendición de cuentas que sólo pueden obtenerse con sistemas de evaluación confiables y pertinentes. Destacaban además los autores que todo ejercicio de evaluación es instructivo de suyo, porque permite que los actores adquieran una mejor comprensión de lo que están haciendo, porque permite resaltar las iniciativas exitosas y explicar el contexto de sus logros, con lo cual puede difundir la capacidad de innovación. Más aun, decían, los hallazgos de las evaluaciones conducen a reconsiderar prioridades y la compatibilidad entre deseos y recursos.

Los documentos más recientes de estrategia sectorial del Banco Mundial, un actor institucional que viene cobrando creciente presencia en el planteamiento de la agenda educativa en nuestra región, también hacen llamados para que las prioridades educativas se determinen cada vez más en sobre la base de estándares de desempeño o *performance* y a la medición de su logro mediante evaluaciones apropiadas, así como también sobre la base del análisis económico de los resultados en el mundo del trabajo.

El tema también ha merecido la atención prioritaria de la Comisión Internacional sobre Educación, Equidad y Competitividad Económica en las Américas del PREAL, constituida por un grupo de líderes políticos y empresariales de diversos países del continente y cuya misión incluye crear una base de apoyo amplia y activa para promover el cambio educativo y desarrollar perspectivas modernas, informadas, sobre política educacional y reforma educativa en América Latina. La primera recomendación de su difundido pronunciamiento **El futuro está en juego**[4] es que los gobiernos fijen estándares para el sistema educativo, que midan los avances hacia el logro de los mismos mediante pruebas nacionales, regionales e incluso internacionales y que utilicen los resultados para revisar sus programas y reasignar sus recursos. Más recientemente, la Comisión Centroamericana para la Reforma Educativa convocada también por el PREAL, recomendó que los países de la subregión establezcan estándares comunes y ampliamente consensuados, así como también un sistema unificado de medición de su logro y la amplia divulgación de sus resultados[5].

También el Plan de Acción de la Segunda Cumbre de las Américas compromete a los gobiernos a establecer sistemas nacionales, subnacionales o subregionales para evaluar la calidad de la educación, midiendo el desempeño de distintos actores educativos, y los efectos de la introducción de innovaciones y otros factores asociados a los logros en el aprendizaje.

---
[3] Delors 1996.
[4] Comisión Internacional 1998:
[5] Comisión Centroamericana 2000: 24 - 25

Por último, tras la evaluación de los diez años del cumplimiento de los objetivos de la conferencia de Jomtien, los países del continente formularon un "Marco de Acción Regional" para los próximos quince años, entre cuyos considerandos y acuerdos se menciona la necesidad de establecer estándares de calidad y de procesos permanentes de monitoreo y evaluación. Se estableció el compromiso de los países de "organizar sistemas apropiados de monitoreo y evaluación que consideren las diferencias individuales y culturales, se basen en estándares de calidad acordados nacional y regionalmente y permitan la participación en estudios internacionales", así como garantizar la participación de la sociedad civil en "la elaboración, monitoreo y evaluación de las políticas educativas…planes y programas…" y "establecer parámetros que identifiquen las responsabilidades de los recursos humanos que actúan en el sistema educativo, así como los mecanismos y políticas de apoyo para la administración del personal"[6].

De ser válidos los supuestos sobre la base de los cuales muchos países y organizaciones vienen fundamentando estas demandas, habría mucho por hacer aún en nuestra región del mundo. Con todo lo avanzado hasta hoy en día, no es posible argumentar que en la mayoría de los países de América Latina existen ya mecanismos permanentes, sólidamente institucionalizados, de establecimiento y revisión de metas de aprendizaje y de medición de su logro, que estén claramente enmarcados en políticas y normas respecto al uso que se quiere y puede hacer con la información resultante y que sean parte también de un conjunto de políticas claras de responsabilización por los resultados del sistema escolar. Como es el caso de muchos de los otros elementos de las reformas educativas en marcha, quedaría aún un buen trecho de recorrido antes de poder afirmar que están ya establecidas visiones y metas ampliamente compartidas – no necesariamente consensuales – respecto a la orientación de dichos sistemas de evaluación, o que rige ya una aceptación generalizada de responsabilidades claramente asignadas a todos los actores que intervienen o debieran intervenir en su gestión.

El fomento de un diálogo horizontal, efectivo y permanente, intra- e intersectorial, (la / el "fronesis" a que alude Rosa María Torres) sobre las reformas y programas que deben introducirse en la educación, particularmente aquéllas conducentes a hacer más viables políticas positivas de responsabilización, es uno de los desafíos más importantes que debe confrontarse en la región. Es también uno de los aspectos con frecuencia menos atendidos por buena parte de las reformas en curso y es algo que UNESCO/OREALC, así como otros organismos regionales e internacionales focalizados en el mejoramiento de la educación latinoamericana, pueden y deben promover intensamente.

Es imperativo encontrar maneras de pasar de las consultas muchas veces meramente formales o "de cumplido" realizadas por los Ministerios de Educación y de los cuestionamientos más contestatarios que propositivos por parte de la sociedad civil organizada que frecuentemente han caracterizado muchas de las discusiones de los últimos años a intercambios constructivos de opiniones y propuestas, que comprometan en algún grado a las partes involucradas y que lleven a consensos mínimos sin los cuales ningún intento de reforma podrá tener éxito.

En el área de los llamados *estándares educacionales*, es crucial enfrentar la tarea de definir y construir consensos sobre contenidos esenciales y metas de aprendizaje y desempeño que todos los alumnos deberían alcanzar gracias a su experiencia escolar. Es necesario establecer con claridad y con participación de toda la sociedad qué es lo que los alumnos de distintos ciclos de estudios deben saber y saber hacer, así como cómo puede demostrarse y medirse que realmente han logrado esas metas de aprendizaje. Es necesario además establecer mecanismos que permitan vincular estas metas con todos los demás componentes de las reformas en curso. En cambio, cómo cada escuela, docente o familia intente lograrlas es una materia que debe determinar cada uno de ellos, contando con apoyo y sugerencias provenientes de diversas fuentes. No se trata de uniformizar la pedagogía, sino de asegurar los resultados y la búsqueda de logros siempre mayores y mejores.

---

[6] UNESCO, PNUD, FNUAP (2000).

En el área de *evaluación de logros* es también perentorio avanzar bastante más, tanto en la que debe realizarse dentro del aula y la escuela, para permitirle al docente y los directivos adecuar sus prácticas pedagógicas para hacerlas más efectivas según los estilos y ritmos de aprendizaje de sus alumnos, como en aquella evaluación externa de resultados, que permita la evaluación de la efectividad del sistema escolar en su conjunto. Esta última puede permitir una identificación más generalizable de los factores escolares y extra-escolares que inciden en el desempeño y una necesaria retroalimentación a los otros componentes del sistema (currículo, textos, capacitación y formación docentes, políticas de gestión, etc.). Es importante adaptar las formas de las mediciones a los fines que se persigue, fines que deben ser discutidos públicamente con el objetivo de maximizar el consenso al respecto y que deben ser revisados periódicamente, a la luz de los avances y logros de etapas previas y de los cambios introducidos en políticas, planes y programas. Es igualmente importante seleccionar las técnicas que mejor se adapten a esos fines y desarrollar las capacidades profesionales necesarias para aplicarlas.

Por último, puede señalarse que los estándares educacionales pueden y deben servir no sólo para alinear las evaluaciones y los otros componentes del sistema educativo, sino también para propiciar *prácticas de responsabilización por los procesos y los resultados*. Los logros y las brechas entre éstos y las metas deben poder ser analizados y ligados a cada uno de los diversos responsables de cuánto y cuán bien aprenden los estudiantes. Ellos incluyen no sólo a los docentes que les han enseñado, sino a quienes desarrollan y adaptan el currículo, a los productores de textos, a las autoridades locales y regionales, a los mismos padres de familia y a los propios estudiantes. Se trata ésta de una tarea compleja y delicada, que requiere un perfeccionamiento de capacidades especiales que hay que promover activamente.

2. La situación actual

La medición de logros de aprendizaje

Es importante destacar y celebrar que a lo largo del decenio pasado casi todos los países latinoamericanos han estado haciendo grandes esfuerzos para medir la calidad del aprendizaje escolar mediante sistemas de pruebas nacionales. En muchos de ellos se realizaron incluso anteriormente algunas mediciones estandarizadas más o menos masivas, ya sea para evaluar el impacto de programas o proyectos o como parte de alguna investigación. Sin embargo, fue recién en la segunda mitad de los noventa que la mayoría de ellos, frecuentemente con apoyo de organismos internacionales, se decidió a aplicar diversas herramientas que le permitirían medir y evaluar sistemáticamente los aprendizajes, con el objetivo de proveer información al sistema educativo y a la sociedad que podría servir como un insumo para tomar decisiones y mejorar los procesos educativos.

En casi todos los países de la región se han realizado ya mediciones nacionales o estatales de logros para alumnos de varios grados escolares y en una diversidad de áreas o competencias curriculares. En muchos casos, hay ya más de una experiencia de aplicación y se ha publicado y difundido al menos parte de sus resultados. En algunos países se han realizado también estudios para investigar qué factores escolares y extraescolares son los que más fuertemente inciden sobre esos logros, con la intención de utilizar los hallazgos para mejorar el diseño e impacto de programas.

Pese a estos grandes esfuerzos, y aun cuando en la mayoría de los casos se cuenta con un sistema inicial de pruebas estandarizadas e incluso de desempeño, como se mencionó anteriormente, en muchos casos aún no puede decirse que existe ya un mecanismo permanente, sólidamente institucionalizado, de medición de la calidad. Tampoco se cuenta, en la mayoría de los casos, con instrumentos que permitan hacer seguimiento a lo largo del tiempo de los resultados de aprendizaje, que permitan vincularlos de manera válida y confiable a los diversos factores contextuales y los propiamente escolares que inciden sobre

ellos y/o hacer un seguimiento a lo largo del tiempo de los resultados e impactos de distintas políticas o programas introducidos durante el período en cuestión.

Muchos de estos esfuerzos no han sido lo suficientemente sistemáticos ni sostenidos. Ha habido falta de continuidad del trabajo, agudizado durante periodos de cambios de gobierno, cada uno de los cuales trae una perspectiva diferente sobre la importancia de la medición de la calidad. Esto se traduce en falta de actualización de las pruebas, una tendencia a ignorar sus resultados y a no usarlos adecuadamente en el proceso de toma de decisiones por las autoridades o por los usuarios del sistema. Obviamente, esto lleva también a un insuficiente impacto de las evaluaciones sobre el conjunto del sistema educativo.

En muchos países, los fines, objetivos, los enfoques y los procedimientos seguidos por los sistemas no han sido discutidos ni entendidos suficientemente por todos los grupos sociales que deberían constituirse en sus usuarios. En algunos casos, ni siquiera han sido discutidos ni aceptados por los actores del mismo sector educativo, por no mencionar a la totalidad de los funcionarios de los niveles centrales donde se formulan y supervisan las políticas y sus resultados! Es justamente debido a esta falta de definiciones básicas y de compromisos sociales claros con la calidad y transparencia de los sistemas, que éstos son aún muy vulnerables a cambios en la administración del gobierno e incluso a cambios en la identidad de los responsables de un ministerio.

Sobre la base de Ravela et al. (1999) y observaciones adicionales, podrían identificarse cerca de una veintena de fines diferentes que persiguen o podrían perseguir los sistemas nacionales de evaluación de logros de aprendizaje:

Finalidades diversas que pueden perseguir los sistemas de evaluación de logros de aprendizaje escolar

1. Informar a la opinión pública y generar una cultura social de la evaluación.
2. Contribuir a la generación de conocimiento, proporcionando insumos para la investigación aplicada sobre el funcionamiento de los sistemas educativos, las prácticas de enseñanza, el impacto de las variables sociales sobre el aprendizaje de los niños (también identificar los factores determinantes del logro) y los tipos de intervenciones más efectivos para mejorar los aprendizajes.
3. Construir un “mapa de situación” del sistema educativo con el fin de identificar áreas o unidades prioritarias de intervención y tipos de intervenciones necesarias, de manera de garantizar igualdad de oportunidades de aprendizaje y focalizar consecuentemente los recursos.
4. Comunicar valores y expectativas de logros.
5. Motivar mejoras y logros vía la comparación, competencia o emulación.
6. Alertar a la opinión pública sobre deficiencias en los resultados de los sistemas educativos y la necesidad de que se apoye intervenciones para su mejoramiento.
7. Identificar metas claras, mensurables y comunicables, centradas en los aprendizajes, para los esfuerzos de mejoramiento del sistema educativo, que faciliten la movilización y apoyo de la opinión pública y otros responsables.
8. Devolver información a las escuelas y maestros para que éstos examinen en detalle los resultados de su trabajo y mejoren sus prácticas pedagógicas
9. Brindar a los padres de familia información que les permita evaluar y controlar la calidad de las escuelas.

10. Contribuir a establecer (o monitorear logro de) estándares de calidad para el sistema educativo.
11. Certificar el dominio de un núcleo de conocimientos y capacidades o competencias por parte de los alumnos que finalizan un determinado nivel de enseñanza.
12. Seleccionar u ordenar a los estudiantes (o a escuelas o jurisdicciones) para acreditar su elegibilidad para ciertos privilegios, acceso a programas o, incluso, sanciones.
13. Evaluar el impacto de políticas, innovaciones o programas específicos.
14. Retroalimentar el currículo y los planes de estudio.
15. Realizar estudios de tipo costo-beneficio, que orienten las decisiones para una distribución más eficiente y efectiva de recursos siempre escasos.
16. Contar con argumentos persuasivos para obtener mayores recursos del presupuesto público o proponer cambios en orientaciones generales del sistema educativo.
17. Demostrar el peso del factor docente en los aprendizajes e influir en la formación, capacitación y desempeño de los maestros.
18. Evaluar la productividad de los maestros a los efectos de establecer un sistema adecuado de incentivos.
19. Promover una responsabilización efectiva de todos o algunos de los actores involucrados en los procesos educativos, incluyendo a los propios alumnos.

Sostienen Ravela et al.[7]:

> *… es imposible que un mismo diseño del sistema de evaluación sirva para todos los fines señalados… Algunos pueden ser perseguidos con un mismo diseño, pero otros son incompatibles entre sí…. cada una de estas finalidades tiene sus propias exigencias técnicas. Para algunos de estos fines son adecuados ciertos tipos de pruebas que no lo son para otros. Para algunos de estos fines se requieren ciertos tipos de muestras que no son adecuadas para otros…Las definiciones técnicas que sirven para un fin no sirven para otro o, lo que es peor, pueden dar lugar a graves malentendidos cuando se las utiliza para ese otro fin.*
>
> *La experiencia indica que en muchos países de la región ha sido insuficiente la reflexión acerca de los fines específicos que se espera cumplan los sistemas de evaluación dentro de un país, así como acerca de las definiciones técnicas más adecuadas para cada fin. Muchos países han trabajado a partir de un propósito general de informar sobre los resultados del sistema educativo para contribuir a su mejoramiento, pero sin diseñar una estrategia más específica. Por otra parte, es bastante común que sobre la marcha las autoridades ministeriales comiencen a demandar que las evaluaciones sirvan para nuevos propósitos o que aporten información para fines para los que no fueron diseñadas.*

La mayor parte de países no han elaborado ni se han comprometido con un plan de trabajo detallado de largo plazo respecto al desarrollo de sus capacidades de evaluación ni a la revisión periódica de sus objetivos, de manera de poder evaluar y de ser necesario retrazar sus estrategias para adaptarse a posibles cambios en sus finalidades prioritarias. El costo ya incurrido en el desarrollo de ciertas capacidades y enfoques técnicos los induce a una cierta inercia y se continúa aplicando métodos de medición que ya no están en capacidad de dar respuesta a las nuevas preguntas y necesidades que van emergiendo.

---

[7] 1999: 8

Se reconoce -- con mucha reserva, y no en todos los países -- las debilidades técnicas que afectan los procesos de desarrollo y validación de los distintos instrumentos de medición. Pero se prefiere no discutirlos públicamente por temor a que ello debilite o deslegitime el proceso necesariamente gradual de desarrollo de capacidades técnicas y políticas para consolidar la evaluación. Como en muchos otros terrenos de acción educativa, aquí tampoco parecería existir el derecho a experimentar ni mucho menos a equivocarse, aun cuando, en la mayoría de los casos, no se han atado "consecuencias costosas" o "implicancias altas" (*high stakes)* a los resultados obtenidos por alumnos o escuelas.

En parte debido a esta cerrazón al escrutinio externo, no siempre se aplican procedimientos que garanticen que las pruebas sean instrumentos válidos para los usos que se les pretende dar ni se han instalado mecanismos institucionales que aseguren que la calidad y pertenencia de las mismas irá mejorando con el tiempo.

Por ejemplo, la idea de que la validez de una evaluación depende fuertemente del uso que se pretende dar a los datos, es decir, de las acciones, decisiones e inferencias que se realizarán sobre la base de las mediciones, y que no es sólo función de la "calidad intrínseca" de los instrumentos de medición ("que efectivamente midan lo que se quiere medir", la concepción tradicional y ya caduca de la validez), no está del todo asentada en nuestro medio. Tampoco lo está la idea de que la validez es una cuestión relativa, de grado. No hay mediciones perfectamente válidas, sino que hay unas más válidas que otras, y ello depende del uso que se pretenda hacer de ellas. Y para que los usuarios puedan evaluar la validez de las mediciones para sus propios propósitos, resulta muy importante documentar rigurosamente todos los pasos seguidos en la construcción de las pruebas, cosa que frecuentemente no se ha hecho[8].

Asimismo, aunque todos los sistemas pretenden alejarse de la medición de capacidades más básicas de pensamiento y evaluar la capacidad de aplicación de razonamientos más complejos, existen abundantes indicios de que los instrumentos tienden a seguir enfatizando lo primero: la recordación de hechos, la retención de datos o las definiciones formales de ciertos conceptos, las habilidades para el cómputo o para resolver "problemas" que casi no requieren de procesamiento complejo de información y que se parecen poco a los de la vida real, etc. Es más, con demasiada frecuencia se mide lo que es más fácil medir y luego explicar, no lo que según el currículo es lo prioritario. Salvo excepciones, la interdisciplinariedad que tanto se reclama para la enseñanza no se hace manifiesta en las pruebas mismas.

Otra debilidad frecuentemente observada es que, pese a manifestaciones explícitas de que se pretende utilizar los resultados para retroalimentar el currículo y los planes de estudio, a menudo el tipo de cobertura de áreas de conocimientos (número reducido de ítemes) y el tipo de diseño de las pruebas (referidas a normas), así como el tipo de pruebas (principalmente de opción múltiple) y la calidad de su elaboración (construcción de distractores, por ejemplo) no son adecuados para llegar a entender realmente qué aspectos del currículo son aprendidos exitosamente o para formular recomendaciones sobre posibles énfasis, secuencias o maneras en que podrían ser mejor enseñados.

Por escasez de recursos humanos y financieros en las unidades responsables de la medición (y muchas veces en el país), y por la reserva con que se suele manejar las bases de datos resultantes, los análisis a los que se someten los datos son bastante escasos. Esto difiere de lo que se puede atestiguar cuando, por ejemplo, se revisan las revistas profesionales y académicas de países desarrollados, en las cuales aparecen numerosos artículos que se han basado en investigaciones y análisis de resultados de pruebas de logros para explorar la validez de diversas hipótesis sobre efectos de determinadas variables sobre el rendimiento escolar en distintos contextos. En América Latina, los datos no son normalmente asequibles a investigadores externos a las unidades de medición datos, presumiblemente por temor a los costos políticos de una divulgación no controlada de resultados que se alejan de lo deseable, pero también por los considerables costos que supone poner esas bases de datos en un

[8] Valverde, en Ravela et al. 1999, cap. II.

formato y con documentación de apoyo que permita a posibles usuarios externos manipularlos de manera adecuada. El hecho de que la mayoría de las agencias responsables de las mediciones estén insertas en las estructuras de los ministerios de Educación, y no, como en la mayoría de los países desarrollados, en instituciones no gubernamentales contratadas por aquéllos, refuerza, desafortunadamente, esta tendencia.

La mayor parte de los sistemas de evaluación de la región, junto con la aplicación de las pruebas, recogen un importante volumen de información de carácter contextual, tanto con relación a las características socioculturales de los alumnos, como con relación a las características propias de las escuelas y maestros. El objetivo de esta recolección de datos es medir el grado de determinación o "asociación" entre factores posiblemente explicativos de los niveles de logro encontrados. Sin embargo, este tipo de información es escasamente difundida y poco utilizada en el análisis de los resultados de aprendizaje. Salvo excepciones, la información recogida se emplea tan sólo para establecer clasificaciones elementales de escuelas o alumnos, con el fin de que los lectores de los informes puedan contextualizar sus propios análisis de los datos que pueda realizar un lector de los informes.

Otra debilidad básica de muchos de los emergentes sistemas es que pocos de ellos han establecido pública y claramente, y con anterioridad a sus primeras aplicaciones de pruebas, qué tipo de decisiones se espera poder tomar con los resultados de las mediciones. No se ha estimulado una discusión nacional o al interior de los centros escolares o ámbitos administrativos específicos sobre las implicancias de los mismos.

La carencia de una cultura de rendición de cuentas se evidencia en cierta reticencia o desconfianza hacia las mediciones y en temores a revelar tanto los resultados como la metodología empleada. La debilidad de las estrategias de difusión de resultados y la discontinuidad en su implementación se deriva en algunos casos de las cuestiones anteriormente mencionadas; en otros, se debe a la escasez de recursos humanos para poder confrontar una posible avalancha de cuestionamientos que podrían ser legítimos pero también posiblemente oportunistas y motivados por intereses políticos estrechos. Ello ha llevado a que la diseminación haya sido muchas veces limitada y su utilidad potencial haya sido frenada. También la falta de criterios claros y ampliamente compartidos (estándares de logros, criterios para establecer "líneas de corte" en cuanto a niveles aceptables o satisfactorios de logros, etc.) contra los cuales se deberían contrastar los resultados ha inhibido una actitud diseminadora más asertiva por parte de los organismos responsables en muchos países. No resulta sorprendente, por lo tanto, que, salvo excepciones, las pruebas nacionales de logros escolares parecen hasta el momento haber servido poco para informar las decisiones de políticas educativas y curriculares.

De otro lado, el impacto que los sistemas de evaluación puedan tener en el mejoramiento de la educación depende también del grado en que logren producir y difundir los resultados de manera apropiada, comprensible y relevante para diversos públicos. De ello dependerá también que los sistemas de evaluación logren el sostén político y financiero necesario para su institucionalización. Hasta el momento, en la mayoría de los países no se ha evaluado aún cuán efectivas son sus maneras de difundir información, cómo están siendo interpretados los resultados por diversos sectores clave de la opinión pública ni por actores del mismo sector educativo, ni qué tipo de inferencias hacen éstos a partir de las diversas publicaciones y espacios de debate que puedan existir al respecto. Es presumible que, para muchos padres y maestros, por ejemplo, los gráficos con que se pretende comunicar ciertas distribuciones de resultados, escapen a su comprensión. Se desconoce cómo mejor preparar al periodismo para difundir de manera útil y constructiva los resultados de las evaluaciones – aunque es claro que nada puede garantizar que luego la cobertura que se dé a las mediciones resulte sesgada por intereses políticos más que "de políticas". Es por ello que se hace necesario producir y sistematizar conocimiento acerca de los modos de organizar y presentar la información derivada de las evaluaciones nacionales de logro educativo en la región, sus fortalezas y debilidades.

Participación en pruebas internacionales

La participación latinoamericana en experiencias internacionales de medición de logros de aprendizaje es muy limitada. En 1997, con el respaldo y coordinación de la Oficina Regional de la UNESCO se realizó una evaluación en Amércia Latina del aprendizaje en matemática y lenguaje en niños de tercero y cuarto grados. Participaron 13 países, aunque uno de ellos impidió la publicación de sus resultados y otro no procedió oportunamente a la entrega de las bases de datos, aparentemente, por problemas generados en cambios administrativos.

A la fecha pocos países de la región han participado en otras pruebas internacionales de medición del rendimiento, como puede apreciarse en la tabla siguiente:

Tabla 1. Países americanos participantes en diversos proyectos internacionales de evaluación de la educación

	OREALC	PISA	TIMSS(-R)	RLS-PIRLS	CES
Argentina	X		X	X	
Bolivia	X				
Brasil	X	X			
Canadá		X	X	X	
Colombia	X		X	X	X
Costa Rica	X				
Cuba	X				
Chile	X		X		
Ecuador					
El Salvador					
EE. UU.		X	X	X	X
Guatemala					
Honduras	X				
México	X	X			
Nicaragua					
Panamá					
Paraguay	X				
Perú	X				
República Dominicana	X				
Trinidad-Tobago				X	
Uruguay					
Venezuela	X			X	

Fuente: Tiana 2000 y otras comunicaciones

OREALC: Prueba del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa de la Oficina Regional de la UNESCO, en matemáticas y lenguaje; PISA: Programme for International Student Assessment de la OECD, en matemáticas, ciencias y lenguaje; TIMSS, TIMSS-R: International Trends in Mathematics and Science Study (antes Tercer Estudio Internacional de Matemáticas y Ciencias) de la IEA; RLS, PIRLS: Prueba Internacional de Comprensión Lectora de la IEA; CES: Estudio sobre Educación Cívica de la IEA

Cuando han participado, los resultados han sido generalmente muy insatisfactorios, aun si se restringen las comparaciones a países de similares niveles de desarrollo e ingresos nacionales. En más de un caso, los países se han abstenido a último momento de aplicar las pruebas o de autorizar la publicación de sus resultados. Colombia aprovechó positivamente sus pobres resultados en TIMSS (sólo 4% de sus alumnos participantes logró "clasificarse" dentro del 50% superior en el grado en el cual se inscribió Colombia para las pruebas) para poner en marcha un debate nacional al respecto y plantear una reforma curricular, mejoras en sus textos y en la capacitación docente, y ha decidido participar en PIRLS en el año 2001. Salvo ese caso nacional, se desconoce mayormente el impacto interno que ha tenido la participación de nuestros países en estas pruebas internacionales o los planes de difusión y utilización que tienen los países que recientemente se han incorporado a varios de estos sistemas de evaluación, como Chile, Argentina, Brasil y México.

Lo cierto es que, hasta la fecha, las experiencias de medición no permiten comparar los logros específicos de los países con estándares regionales o internacionales.

Recientemente, han surgido muchas recomendaciones provenientes de organismos internacionales y regionales en el sentido de que es conveniente que los países latinoamericanos participen en esas pruebas internacionales de logro académico. Esas recomendaciones suelen sustentarse en el argumento de que los datos comparativos son muy útiles para informar a los países de su posición relativa y su competitividad vis-a-vis sus

competidores en el mercado global de bienes y servicios. Están también basadas en el argumento de que las pruebas internacionales garantizan más alta calidad técnica y eficiencia que los sistemas nacionales y que pueden contribuir significativamente al desarrollo de capacidades locales en el campo de la evaluación.

Sin embargo, se sabe relativamente poco sobre los beneficios y riesgos potenciales de utilizar estas pruebas internacionales para medir y evaluar la calidad educativa de un país. No se cuenta con suficiente evidencia empírica sobre el impacto real que esas pruebas han tenido en los países que han participado anteriormente en ellas, particularmente en los países en desarrollo. ¿Qué usos específicos se ha dado a sus resultados para mejorar la calidad de las escuelas? ¿En qué medida los participantes han logrado colocar sus inquietudes y sus preguntas fundamentales en la agenda de las evaluaciones? ¿En qué medida las competencias o contenidos que las pruebas miden reflejan objetivos fundamentales de los diversos participantes, y no sólo las de los técnicos que lideran su elaboración? ¿Han servido realmente para el desarrollo de capacidades nacionales que han continuado al servicio del desarrollo de mejores políticas domésticas? Incluso asumiendo que la tendencia hacia una mayor participación se hará "irresistible", ¿cómo pueden los países de la región prepararse técnica y políticamente para una experiencia más beneficiosa?

Son todas estas preguntas que bien valdría la pena intentar responder, aunque fuera tentativamente, antes de tomar decisiones al respecto, particularmente si la incursión en los sistemas internacionales de evaluación de aprendizajes puede acarrear la asignación prioritaria de recursos escasos a ellos, en detrimento de los esfuerzos locales, potencialmente más fácilmente vinculables a necesidades de información más inmediatas.

Estándares o criterios e indicadores para la evaluación de logros

Debe resaltarse también el hecho de que, salvo casos excepcionales, en los países de América Latina no se ha especificado, discutido y difundido cuáles son las expectativas de logro, ni cómo se habría de saber si un examinado ha aprendido lo que debe aprender. Esto, es decir, el hecho de no contar con criterios e indicadores de logros o *estándares*, y el hecho de que muchas de las pruebas nacionales siguen estando referidas a normas (es decir, sólo sirven o debieran servir para comparar el logro de distintos grupos, y, para analizar el impacto de diversos factores sobre el nivel de logros de esos distintos grupos, pero no para determinar cuánto del currículo intencional o del currículo implementado está siendo efectivamente dominado o aprendido por los escolares, ni cuán bien lo están haciendo) – o tienen un tipo "híbrido" del cual desconfían muchos especialistas -- hacen difícil la interpretación correcta de los resultados de muchas de las pruebas de rendimiento escolar.

Reporte tras reporte o artículo de opinión pública informa que los alumnos de un país sólo dominan algo menos o algo más que la mitad de los conocimientos, capacidades, habilidades, destrezas o competencias que se esperaría hubieran desarrollado en un campo disciplinario general una vez alcanzado determinado nivel de su escolaridad, o que los alumnos de las escuelas públicas dominan incluso menos que esa mitad. Esa inferencia es, con frecuencia, ilegítima. La conclusión está predeterminada por la manera en que se construyen las pruebas referidas a normas: eliminando preguntas que en la etapa piloto resultaban muy difíciles o muy fáciles y que reducían la variabilidad en las respuestas, con lo cual no se podían hacer los análisis estadísticos acostumbrados para ese tipo de medición. Ni la cobertura del currículo, ni, con frecuencia, la cobertura de las muestras de estudiantes que son evaluados con ese tipo de pruebas permiten llegar válidamente a ese tipo de conclusión.

En lugar de poder decirnos solamente quiénes están mejores o peores que otros, que es casi todo lo que pruebas referidas a normas pueden contribuir a establecer, buenas pruebas elaboradas con referencia a criterios y estándares permitirían medir cuánto de lo que la sociedad espera que los escolares aprendan está efectivamente siendo aprendido, así como cuán bien está siendo aprendido. Idealmente, la definición de esas expectativas de logro o estándares de desempeño y niveles de ejecución debería hacerse con suficiente claridad y

antelación al diseño y aplicación de las pruebas y con amplia participación de distintos grupos de actores sociales.

Los estándares educacionales definen las competencias y habilidades básicas que deberían adquirir los niños y jóvenes durante su tránsito por el sistema escolar. Definen y comunican con claridad a distintas audiencias, no sólo las especializadas, lo que los estudiantes deberían saber en cada área de conocimientos en cada grado o nivel y lo que deberían saber hacer con esos conocimientos, además de establecer cuán bien deberían poder hacerlo, conforme a una gradiente de desempeños basada tanto en lo deseable como en lo ya observable.

Se piensa que definir con claridad y elevar esos estándares contribuirá a mejorar los logros de aprendizaje de todos los alumnos, movilizando sus propios esfuerzos y los de todos los otros actores del sistema que inciden sobre sus logros. Sería más fácil que se dé ese tipo de movilización si tanto los estudiantes como sus padres y sus docentes supieran que se espera que, hacia el final del cuarto grado, por ejemplo, los alumnos sepan sumar y restar fracciones y sepan como aplicar esas operaciones para la resolución de problemas prácticos de su vida cotidiana, que sepan dar a otros instrucciones precisas para llegar a algún lugar conocido, sepan las diversas razones que indujeron a las colonias americanas a rebelarse contra la autoridad de sus gobernantes europeos y sus representantes locales y puedan expresar en sus propias palabras qué ganaron y qué perdieron ciertos grupos sociales con el cambio habido tras su independencia.

Las evaluaciones de rendimiento serían más fáciles de elaborar, tendrían más sentido y gozarían de mayor legitimidad, si esos estándares fueran ampliamente conocidos y discutidos antes de su establecimiento "formal", en los distintos ámbitos y niveles de gestión educativa. Es más, no se trata tan sólo de que ellos sean conocidos por actores de diversos sectores de la sociedad civil, sino que éstos y sus organizaciones (profesionales, escolares y ciudadanas) deben haber participado o quizás iniciado movimientos en pos de su establecimiento e implementación.

Nadie podría en rigor argumentar que en algún país latinoamericano existe ya un conjunto de estándares altos, estimulantes y legitimados entre amplios sectores de la sociedad o que son el marco de referencia principal para la elaboración del currículo, el diseño de textos y materiales, la formación de docentes y el mismo diseño de su sistema de evaluación.

Dicho lo cual, sin embargo, es conveniente anotar que en casi todos los procesos de reforma curricular en curso en los países de la región se ha expresado explícitamente la necesidad de arribar a consensos sobre qué metas y contenidos son deseables para su población y su sistema educativo[9]. Esas expresiones, sin embargo, muchas veces no han tenido como correlato la implementación de canales efectivos para la generación de dichos consensos. Aun cuando las iniciativas de reforma hayan tenido un origen político, sus procesos de validación se han limitado generalmente a acciones de carácter casi exclusivamente técnico. A diferencia de lo que ocurre con el movimiento en pos del establecimiento de estándares educacionales, particularmente en los Estados Unidos, el proceso de validación curricular y de sus indicadores de logros en América Latina[10] suele darse sólo durante su elaboración inicial, mediante comisiones o grupos de trabajo especiales creados durante el lanzamiento de los programas y hasta el momento de su establecimiento legal o normativo, mientras que "los sindicatos docentes y asociaciones profesionales no se han constituido en agentes permanentes de promoción para la elaboración y renovación de contenidos, lo cual indica que

---

[9] Los párrafos siguientes se basan en el informe de un estudio sobre los contenidos de los documentos curriculares recientemente producidos en el marco de las reformas en los países latinoamericanos, realizado recientemente en el marco de las actividades del Grupo de Trabajo sobre Estándares y Evaluación del PREAL (Ferrer 1999).

[10] Ferrer (1999), refiriéndose al caso de los países de la región con fuerte tradición de planificación curricular centralizada, indica que es en esta fase del diseño curricular donde mejor pueden establecerse estándares de contenido y desempeño.

la validación curricular es uno de los aspectos de las políticas educativas de la región que aún permanecen relativamente centralizados"[11]. La validación y la consulta deberían

> *"consolidarse como proceso permanente, donde la sociedad civil y los diferentes actores del sistema educativo participen en forma dinámica y sostenida para garantizar la renovación y mejoramiento de los contenidos curriculares. Cabe recordar que esta validación es un proceso técnico, en tanto requiere de la participación de especialistas, a la vez que es político y democrático y que por lo tanto requiere de la participación de la sociedad toda en un diálogo donde estén representados los valores e intereses de todos los sectores de la población"*[12].

Sobre este último punto, sin embargo, es necesario tener conciencia de que puede haber un *quid-pro-quo* entre la necesidad y conveniencia de un consenso, tanto de la opinión pública como de la profesional, y la necesidad y conveniencia de la innovación. La necesidad de encontrar convergencias entre todas las partes efectiva o idealmente interesadas, es decir, entre profesores, padres, estudiantes, administradores, profesores universitarios, líderes empresariales y comunales, y diseñadores de políticas puede entrar en tensión, sino conflicto, con el objetivo de lograr estándares óptimos. Un participante en unos de esos esfuerzos para establecer estándares en un estado norteamericano decía:

> *Por un lado, si se suponía que estos estándares debían representar las banderas de la comunidad, entonces tenían que haber reflejado valores y propósitos compartidos. Por otro lado, si se deseaba el cambio, entonces estos estándares tenían que hacer algo más que reflejar las prácticas vigentes. Se necesitaba nuevas ideas, ideas que se alejaban de las prácticas y los supuestos vigentes…*[13]

Debido a lo anterior, los indicadores de logros y competencias curriculares -- dejando de lado cuestiones técnicas que los alejan en mayor o menor grado de las definiciones ideales de lo que son o deben ser los estándares -- carecen con frecuencia de la legitimidad y el consecuente compromiso social necesarios para movilizar y articular los esfuerzos de innovación y reforma tanto en lo pedagógico como en la gestión educativa, y para dar mayor sentido y consistencia a las evaluaciones.

La revisión de los currículos reformados, según también informa Ferrer, revela avances sustantivos respecto a los viejos listados de datos que debían adquirirse en el proceso de la escolaridad, la inclusión de consideraciones reflexivas sobre diversas necesidades de conocimientos y competencias para la vida en una sociedad moderna y compleja, y especificaciones de contenidos mínimos a alcanzar. Pero, ya sea por el aún débil desarrollo de capacidades técnicas en las oficinas curriculares de muchos de nuestros países, por la falta de participación activa y efectiva de expertos externos al sistema escolar, o por el deseo de evitar los conflictos potenciales antes indicados, muchos de estos documentos curriculares no proveen aún parámetros de desempeño y rendimiento que puedan guiar con suficiente claridad los esfuerzos de evaluación. Tampoco parecen contener algo que es hasta más importante, metas altas y crecientes de competitividad y excelencia académicas (términos ya de por sí conflictivos en nuestros sistemas educacionales), comparables a las que explícitamente están persiguiendo las escuelas a las que asisten las elites nacionales y regionales y las que buscan alcanzar los sistemas educativos de otras regiones del mundo.

Desde esta perspectiva, debe llamarse la atención hacia la reciente experiencia de los Ministerios de Educación centroamericanos, quienes, convocados por la Secretaría General de la Coordinación Cultural y Educativa Centroamericana (CECC), acaban de publicar su propuesta de estándares nacionales y subregionales para la educación primaria en matemáticas, ciencias y lenguaje. Es indispensable que dicha experiencia se dé a conocer más allá de las bastante reducidas instancias técnicas y administrativas que hasta el momento han sido informadas al respecto. Debe asegurarse que se inicie en Centro América amplias

[11] Ferrer 1999.
[12] ibid: p. 37.
[13] Ball 1992, citado en CPRE 2000: 4.

discusiones nacionales, regionales y locales sobre los aprendizajes que las escuelas deben lograr que desarrollen sus alumnos y que el seguimiento y evaluación del proceso seguido para su desarrollo y eventual implementación pueda ser aprovechado por otros países y subregiones del continente.

3. Elementos (mínimos)de los escenarios deseables

Un movimiento hacia modelos operativos de sistemas escolares más efectivos requiere el establecimiento de consensos sociales amplios sobre los contenidos esenciales y las metas de aprendizaje y desempeño que **todos** los alumnos de un país, de una región del mismo, de un espacio geográfico determinado, o de un centro escolar en particular, deberían alcanzar gracias a su experiencia escolar. Requiere, por lo tanto, de una comprensión a nivel local de la conveniencia de la adopción de algunos estándares de carácter más universal -- y canales para la participación de actores representativos de la diversidad de situaciones nacionales y locales en el establecimiento de esos estándares y de sus indicadores de logros -- a la vez que requiere de espacios para la inserción de expectativas y necesidades locales.

Un escenario deseable se caracterizaría por un intenso debate local, con participación de actores diversos que incluyen tanto a docentes como a otros líderes sociales, sobre el significado de esos estándares y por la elaboración de indicadores de logros basados en visiones realistas de las brechas de aprendizajes que distinguen los resultados de las escuelas "realmente existentes" de ideales más universales. Las escuelas, las localidades y los sistemas nacionales contarían con "estándares propios", elaborados a la luz de miradas hacia fuera y hacia dentro de sus propios contextos y necesidades inmediatas.

En cualquier escenario deseable se habría establecido con claridad y con participación de toda la sociedad qué es lo que los alumnos de distintos ciclos de estudios deben saber y saber hacer (estándares), y cómo puede demostrarse y medirse que realmente han logrado esas metas de aprendizaje. Cómo es que cada escuela, docente o familia intenta lograrlas sería materia que debe determinar cada uno de ellos, pues no se trata de uniformizar la pedagogía, sino de asegurar resultados y búsqueda de logros siempre mayores.

Los estándares tienen la particularidad de "enganchar" bien con la mentalidad empresarial cuyas organizaciones se han convertido en muchos estados y localidades norteamericanas, por ejemplo, en los más vocales promotores de su establecimiento y utilización como ejes articuladores de los otros componentes de las reformas educativas. Esas organizaciones, así como otros grupos profesionales especializados, centros de investigación educacional, organizaciones gremiales de docentes, instituciones no gubernamentales de promoción de mejoramiento educativo, y diversas alianzas de todos los anteriores tomarían la iniciativa en la propuesta y formulación de estándares y métodos de evaluación, no limitándose a la toma de posición frente a alguna propuesta formulada desde las instancias de gobierno. Los centros escolares o instancias locales de gobierno procederían autónomamente a adoptar alguna/s de esas propuestas.

En un escenario deseable, debería fomentarse la libre participación en sistemas variados de evaluación, nacionales e inernacionales, con atención a ciertas consideraciones sobre usos alternativos de recursos incluyendo el tiempo de los estudiantes y los docentes, tanto para la aplicación de pruebas como para la participación en el análisis y utilización de sus resultados. Habría múltiples (y parsimoniosas) formas de evaluación, y las formas de medición se adaptarían a los fines que se persigue en cada instancia, fines que deben ser discutidos públicamente para maximizar el consenso al respecto.

Se habrá avanzado bastante más en la evaluación de logros, tanto aquélla que debe realizarse dentro del aula y la escuela, que permite a los docentes afinar sus prácticas pedagógicas para hacerlas más efectivas y adecuadas a los distintos estilos y ritmos de aprendizaje de sus alumnos, como en la evaluación externa de resultados, que permite juzgar la efectividad del

sistema escolar en su conjunto, la identificación de factores escolares y extra-escolares que inciden en el desempeño, y la retroalimentación a los otros componentes del sistema (currículo, textos, capacitación y formación docentes, políticas de gestión, etc.).

Se retroalimentará de manera que los resultados guíen las mejoras en la instrucción (identificación de distribuciones "típicas"de niveles de desarrollo de las comprensiones, capacidades y habilidades de los alumnos y propuestas de actividades de aprendizaje que más convendrían desarrollar) y la toma de decisiones por distintos actores (padres, estudiantes, docentes, administradores, funcionarios, diseñadores de políticas). Las formas de utilización de resultados de las pruebas de logros mostrarán un balance entre objetivos de "responsabilización" (accountability) y de mejoramiento escolar

Los sistemas deseables de evaluación otorgarían la debida consideración a aspectos discriminativos que podrían esconderse en sus instrumentos de medición, evitando sesgos étnicos, culturales, regionales y de género que podrían tornar no comparables los resultados obtenidos por diversos grupos de estudiantes.

En un escenario deseable, los estándares y las evaluaciones estarán sirviendo para fomentar prácticas de responsabilización por los resultados. Los resultados en términos de logros estarían siendo constantemente analizados por una variedad de investigadores y analistas de políticas, y ligados con el debido cuidado técnico y político a cada uno de los diversos responsables de cuánto y cuán bien aprenden los estudiantes. Esto incluye no sólo a sus maestros, sino a quienes desarrollan y adaptan el currículo, a los productores de textos, a las autoridades locales y regionales, a los mismos padres de familia y a los propios estudiantes.

Una función importante de las evaluaciones y de la difusión de sus resultados sería el orientar y responsabilizar a los políticos por sus decisiones o falta de decisiones respecto a las orientaciones generales y las políticas y programas específicos que promueven los gobiernos nacionales.

Se habrían diseñado y legitimado arreglos organizacionales alternativos que permitan un más rápido y profundo proceso de institucionalización y legitimación de las evaluaciones, como podría ser quizás la delegación de la elaboración, administración y/o análisis de las pruebas a organismos privados, o la constitución de organismos con niveles mayores de autonomía respecto a la autoridad política de turno.

Existiría un nivel razonable de participación de los países de la región en pruebas internacionales, cuyos resultados son usualmente más impactantes sobre prensa y la opinión pública que las pruebas nacionales. Los resultados, previsiblemente insatisfactorios durante un buen tiempo más, proveerían a los *policy-makers* con argumentos fuertes para reclamar mayor apoyo presupuestal al sector. Permitirían avanzar también hacia una comprensión má generalizable sobre cómo y en qué condiciones inciden ciertos factores escolares tales como las edades de ingreso al sistema, la longitud de la jornada escolar, los métodos instruccionales, la capacitación docente, etc. sobre el desempeño del propio sistema, así como en el conocimiento sobre las relaciones entre rendimiento estudiantil y factores extra-escolares y entre aquél y otras características de los países (crecimiento, empleo, etc.).

Los productores de pruebas internacionales estarán mas sensibilizados a las necesidades de sus "clientelas" más diversificadas, y, al contar con medios técnicos más sofisticados, podrán acomodar las mediciones a prioridades curriculares diversas, pilotear y probar la validez y confiabilidad de las preguntas y estímulos con que realizarán sus evaluaciones, preguntas y estímulos que se acercarán más a simulaciones de situaciones "de la vida real" y a la medición del desempeño que las pruebas actualmente disponibles. Es posible que las evaluaciones se "inserten" de manera menos disruptiva en la entrega del currículo, particularmente a medida en que se introduzcan nuevas tecnologías para la enseñanza y el aprendizaje, algunas de ellas a distancia pero potencialmente "personalizables".

4. Escenarios probables y sus consecuencias

Las tendencias al incremento de la competitividad entre naciones para atraer inversiones sobre la base de las capacidades productivas de sus recursos humanos generará necesidades y prácticas de "certificación de competencias", en algunos casos, de naturaleza más estrecha que lo que sería deseable como objetivo de los sistemas educativos. Pero, en otros, aun ese ámbito más pequeño de capacidades, habilidades y destrezas, será más amplio que lo que los sistemas están efectivamente "entregando" como oportunidades de aprendizaje a sus alumnos. En ésta, como en otra áreas de la evaluación educacional, es previsible un rol ampliado de pruebas internacionales (o de pruebas producidas por empresas transnacionales), lo que podría, al menos en el corto plazo, acarrear un debilitamiento de las aún débiles capacidades de monitoreo o seguimiento con fines de mejoramiento de los aprendizajes y de evaluación con objetivos de responsabilización de diversos actores. Habría poca capacidad para promover la incorporación en el diseño de esas pruebas de criterios o contenidos basados en contextos realmente locales. No se trata simplemente de adaptar los formatos de preguntas a referentes o lenguaje y expresiones nacionales, sino de la capacidad de proponer prioridades respecto a las competencias y habilidades que realmente requieren mayor atención. Es previsible que se requiera un tiempo considerable para desarrollar esas capacidades – técnicas y políticas – y es conveniente empezar desde ya.

Se pronunciará la tendencia hacia una mayor segmentación de "mercados evaluativos", con consecuencias en lo que se refiere a la equidad en las oportunidades de aprendizaje. Muchas escuelas e instituciones de educación superior privadas de elite se someterán a sistemas internacionales tipo Bachillerato Internacional o las que puedan producir y administrar directamente instituciones como ETS, con criterios o estándares internacionales de logros de aprendizaje, mientras que la mayoría de las escuelas públicas serían sujetos de evaluaciones referidas a criterios más localistas y de menor nivel de expectativas o exigencias.

Las líneas que actualmente separan el currículo (los currículos, en realidad) de las evaluaciones externas, y que separan también las evaluaciones externas de la evaluación interna formativa y sumativa que se realiza día a día en las aulas, se borrarán – como también se irán atenuando las líneas divisorias entre lo que se aprenderá en el aula y aquello que se aprehende en muchos otros contextos educativos. En un inicio, esto puede debilitar la calidad y la relevancia de los procesos educativos escolares, lo que debe ser objeto de monitoreo y medidas preventivas.

La diversificación de proveedores educativos, que trascenderán largamente a las escuelas formales de los sistemas, obligará a establecer mecanismos reguladores, algunos de ellos vía información (acreditación nacional o internacional de instituciones y sus resultados), otros vía incentivos o restricciones a su funcionamiento.

En los procesos de diseño y rediseño de los sistemas nacionales de evaluación y de las pruebas nacionales de logros, se incrementarán las tensiones entre las funciones de responsabilización y las de promoción del mejoramiento de los aprendizajes. Dada la escasez de recursos de todo tipo, habrá presiones para que un único conjunto de pruebas sirva para ambos propósitos, y, posiblemente, bastantes más entre los mencionados al inicio de este texto. Esto podría generar resultados e inferencias no válidos y potencialmente costosos en términos de contribuir argumentos o justificaciones para decisiones equivocadas, y hace conveniente el establecimiento de "observatorios" externos a los sistemas que alerten sobre esos riesgos y eduquen a los usuarios respecto a los usos adecuados de la información resultante.

En muchos de los países de la región, para fortalecer el proceso de construcción y legitimación social de los sistemas de evaluación de logros escolares se ha subrayado -- al menos en el discurso -- sus funciones de retroalimentadores del currículo, de insumo para el mejoramiento de la capacitación docente y la reflexión en las instituciones escolares sobre las fortalezas y debilidades de sus planes de estudio y prácticas pedagógicas. Muchas veces se ha descartado explícitamente su utilización para fines tales como la evaluación de estudiante

individuales, su selección para cursos de estudios superiores alternativos, la evaluación de docentes y directivos, etc. Es muy probable – de hecho, ya está ocurriendo en varios países – que se pretenda en los próximos años utilizar los resultados de los alumnos para evaluar al menos parcialmente el desempeño docente. La "amenaza" de este tipo de utilización de las pruebas, así como la posibilidad de que se aten algunas "implicancias altas" *(high stakes)* en lo que se refiere a la graduación, retención de diplomas o retención en ciertos grados, posibilidad de una "segunda certificación" de logros que pueda llegar a tener una mayor valoración en un mercado laboral cada vez más competitivo, etc., generará conflictos que requerirán de capacidades de interlocución entre las unidades responsables de las evaluaciones y la sociedad civil mayores a las que hasta ahora generalmente se han visto obligadas a desplegar. Una posibilidad es que, frente a la posibilidad de esos conflictos, se descarte del todo la ligazón de algún tipo de estímulo o sanción frente a los resultados en las pruebas, lo cual podría debilitar su impacto potencial sobre el mejoramiento escolar.

Por otro lado, la reacción de, por ejemplo, las organizaciones gremiales docentes ante la "amenaza" del uso de las evaluaciones estudiantiles como parte de la evaluación de su propio desempeño profesional -- o ante la inminencia de cualquier forma de evaluación de desempeño percibida como impuesta por instancias burocráticas o gubernamentales inducidas por consideraciones de diversa índole -- . Es poco previsible que en el lapso de los próximos quince años se reconstituya y consolide la identidad profesional de los docentes, muy debilitada por varias décadas de desatención y burocratización de sus carreras. Pese a que esporádicamente emergen visos de reacción ante esta situación, no es aún algo claramente perceptible que las organizaciones del gremio estén encaminadas a definir -- autónoma aunque consultivamente -- los estándares de desempeño que se exigirá a aquéllos que quieran trabajar como maestros ni a definir el tipo de formación inicial y continua, así como una estructura de carrera, que asegurarían que los docentes escolares cuenten con los conocimientos, habilidades e incluso actitudes necesarios para practicar adecuadamente dicha ocupación. Sin esa autodefinición de estándares profesionales (y sin su legitimación ante la sociedad), difícilmente se podrá diseñar un sistema de evaluación de desempeño docente adecuado, y se tenderá a "echar mano" a los datos de las pruebas de rendimiento escolar, de difícil "adscripción" al comportamiento de ningún docente en particular. Esto podría generar mayor rechazo o deslegitimación a la evaluación de logros escolares.

Por supuesto, está el conocido riesgo de "estrechamiento del currículo" y de la "enseñanza orientada a las pruebas", particularmente allí donde se continúen aplicando pruebas referidas a normas y con especificaciones técnicas deficientes (insuficientes cuadernillos rotatorios y cantidad y diversidad de reactivos, por ejemplo), y especialmente si se empieza a ligar algún tipo de implicancias a los puntajes obtenidos por los alumnos o las escuelas.

La aplicación repetida de pruebas, el análisis mejorado de los factores que explican el rendimiento estudiantil y el mayor acceso de actores sociales a la discusión de esos resultados forzarán una revisión continua del currículo y el establecimiento de canales horizontales y verticales tanto dentro del sistema escolar y de puentes de comunicación hacia fuera del mismo, que permitirían hacer de esa revisión curricular (o de estándares) un proceso permanente. Esto implica una ruptura deseable en la "manera de ser y hacer" de los organismos rectores de las políticas educacionales en la mayor parte de nuestros países e inversiones adicionales a las ya considerables realizadas en los últimos años en la reforma curricular, entendida generalmente, pese al discurso, como una experiencia limitada en el tiempo (y también en el espacio, generalmente el burocrático), pero probablemente tendrá que confrontar resistencias de las instancias ministeriales y de los que tienen a su cargo las bolsas del Tesoro Público. Mucho dependerá de la capacidad de los "otros actores" de demandar una actualización permanente y de sugerir maneras no costosas y descentralizadas de realizarla, y resulta necesario fomentar su participación en su diseño.

5. Los próximos pasos

Como propone el Grupo de Trabajo sobre Estándares y Evaluación del PREAL[14] que coordina la autora de este documento, la agenda para los próximos años debe incluir necesaria, integrada y prioritariamente:

> *1. Que en cada país, y en la región, se haga explícita la estrategia a través de la cual se espera que el sistema de evaluación de logros contribuya a la mejoría de los aprendizajes y se genere consenso al respecto.*

Para orientar el desarrollo de la evaluación y maximizar su contribución al logro de los objetivos de las reformas educativas latinoamericanas, ya no basta tener en cuenta su propósito general de hacer más transparentes los resultados de los procesos escolares. La variedad de estrategias utilizadas en los diversos países de la región y en otros lugares para buscar que el sistema de evaluación tenga impacto en el sistema educativo revelan una diversidad de objetivos subyacentes, y no en todos los casos ellas parecen representar las opciones técnicas que mejor se ajustan a los objetivos explícitos de los sistemas de evaluación, revelando tal vez el peso implícito de factores políticos, de corto o mediano plazo[15]. Es posible que esas inconsistencias estén a la base de la lentitud con que en algunos países se está pudiendo utilizar las evaluaciones como instrumento para el cambio educativo.

Es indispensable reflexionar y llegar a decisiones – revisables, por supuesto – sobre las implicancias e incentivos que se atarán al desempeño estudiantil medido por las pruebas nacionales o si se preferirá que cumpla una labor principalmente informativa. Es necesario repensar la vinculación entre las mediciones de rendimiento y los procesos (que deberían ser continuos) de desarrollo curricular o formulación de estándares y los canales que se usarán para asegurar una retroalimentación y complementación positiva de esos y otros procesos de gestión para el mejoramiento de la calidad de la educación.

Vinculado a lo anterior están las decisiones sobre la elaboración de exámenes nacionales de "segunda certificación" de cumplimiento de expectativas para la educación básica u otros niveles o la "atadura" de implicancias a esos exámenes (no promoción al siguiente grado o nivel), o, por ejemplo, la utilización de los resultados de este tipo de pruebas para la acreditación de centros escolares, el establecimiento de incentivos para la carrera y el desarrollo profesional docente.

Dependiendo de los fines y estrategias, se tendrá que determinar la amplitud de la cobertura curricular de las pruebas y de la cobertura (muestras o censos) poblacional de las mediciones – y no sólo en base a los costos. Será necesario también determinar si las pruebas deben ser comparables a lo largo del tiempo, en qué áreas y en qué niveles del sistema, y con qué metodologías se asegurará la validez de esas comparaciones. Asimismo, se deberá reconsiderar la frecuencia de los operativos o aplicaciones nacionales.

Un desafío importante a considerar es de qué modo involucrar en estos debates sobre fines y estrategias a otros actores de la política educativa y no sólo a los técnicos directamente involucrados en las evaluaciones.

> *2. Que se mejore la calidad técnica de diversos elementos de los sistemas, empezando por el diseño de los instrumentos de medición y de recolección de datos complementarios y por los modos de procesar y reportar los resultados.*

Como se argumentó en la segunda sección de este documento, y como han argumentado con énfasis muchos especialistas en el tema, es conveniente tener muy presente que las decisiones técnicas que son aptas para ciertos fines no lo son para otros. Por tanto, es necesario garantizar la congruencia entre las opciones de política y las decisiones técnicas que definen el diseño del sistema de evaluación. Asimismo, es imprescindible mejorar el diseño de los

---

[14] Ver Ravela et al. 1999: 64 -71.

[15] Para un análisis desde esta perspectiva, ver Benveniste 1999.

instrumentos de medición y garantizar una interpretación apropiada –válida- de los resultados que se obtienen. Todo ello exige intensificar los esfuerzos de capacitación de cuadros técnicos y la acumulación de conocimiento y experiencia en una materia que aún es nueva en la región y sobre la que existe escasa "masa crítica".

Considero que en este terreno es importante encontrar maneras de superar ciertos condicionantes políticos que han llevado a una suerte de chauvinismo perceptible hasta en los esfuerzos más exitosos por implantar y lograr aceptación social de los sistemas nacionales de evaluación, y promover un contacto más intenso y productivo entre los profesionales de la región vinculados al área de la evaluación con especialistas de la comunidad internacional. Sería muy conveniente instaurar procesos de "auditoría compartida" de los instrumentos y procedimientos de evaluación. Si bien en la actualidad hay ya un número de consultores internacionales que han tenido la oportunidad de asesorar e informarse sobre el desarrollo y los límites de las capacidades en diversos países, son pocos los profesionales latinoamericanos que han podido hacerlo conjuntamente con ellos y aprender colectivamente en el proceso.

De esos procesos podría emerger la oportunidad de formular estándares técnicos para diversos aspectos de las evaluaciones: construcción y validación de pruebas e ítems, procedimientos de implementación de los operativos de evaluación, procesos de conformación y procesamiento de las bases de datos y reportes y usos de resultados.

3. Que se definan las estrategias de uso y difusión de los resultados de las evaluaciones.

Actualmente, es poco lo que se sabe acerca del uso e impacto que la información producida por los sistemas de evaluación de aprendizaje tiene en sus potenciales usuarios. Se carece de evidencia empírica sólida y suficiente acerca de

- el modo en que los resultados son analizados y utilizados en las escuelas;
- el grado en que las familias y la opinión pública reciben y comprenden la información;
- el modo en que la misma es empleada como insumo en la toma de decisiones de política educativa por parte de los Ministerios de Educación y otros organismos responsables de la misma;
- el grado en que las bases de datos son aprovechadas por académicos y centros de investigación para producir conocimiento.

Es necesario propiciar el desarrollo de trabajos de investigación que permitan recoger esa evidencia, tal cual se lo ha propuesto el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación de la OREALC, así como evidencias sobre los efectos que distintos enfoques y estrategias respecto al reporte y la divulgación de resultados de las evaluaciones nacionales tienen en diversos públicos, como se lo ha propuesto el PREAL.

También sería conveniente establecer foros y eventos que permitan "escuchar a los destinatarios". Es necesario identificar:

- ¿qué tipo de información esperan recibir del sistema de evaluación de aprendizajes las diversas audiencias –periodistas, padres, maestros, políticos, autoridades y técnicos de los Ministerios de Educación-?;
- ¿cómo perciben la información que actualmente se les está entregando?, ¿han podido comprenderla?, ¿la han utilizado de algún modo?;

- ¿qué visión general tienen acerca de los sistemas nacionales de evaluación de aprendizajes?, ¿cuáles son sus expectativas y prejuicios acerca de los mismos?.

Ello permitiría analizar la demanda potencial de información, aprender acerca de los modos pertinentes de informar a los diversos tipos de usuarios potenciales y desarrollar diferentes tipos de formatos de informe adecuados a cada uno de ellos. Dado que la mayoría de los potenciales destinatarios probablemente no tenga una noción cabal de lo que espera de un sistema de evaluación de aprendizajes por tratarse de algo relativamente nuevo en la región, foros de esta naturaleza podrían aportar pistas para desarrollar estrategias de "formación de la demanda", es decir, ayudaría a pensar qué tipo de acciones desarrollar para ir conformando en nuestros países una cultura con relación a la evaluación de aprendizajes que incluya aspectos tales como la conciencia acerca de la necesidad de contar con esta información, los alcances y limitaciones de la misma, el tipo de interpretaciones válidas, los modos de utilizarla y los tipos de información que es posible demandar al sistema de evaluación para fines específicos.

Además de lo anterior, sería conveniente:

4. Reunir evidencia sobre el impacto que a la fecha han tenido la aplicación y difusión de resultados en pruebas internacionales de rendimiento académico en los países participantes, particularmente los países latinoamericanos y otros países en desarrollo.

5. Documentar la experiencia regional en el desarrollo de pruebas referidas a criterios, particularmente en lo que se refiere a la medición de competencias complejas mediante pruebas de desempeño, a la confiabilidad de sus métodos de calificación y a las maneras de reportar de manera comprensible y útil sus resultados.

6. Documentar el impacto que los actuales sistemas nacionales de evaluación externa de logros escolares está teniendo sobre las prácticas pedagógicas y evaluativas en las aulas, tanto aquéllos efectos deseables como los indeseables desde la perspectiva de los fines explícitos de los mismos sistemas de evalución, así como desde las emergentes perspectivas de "la buena pedagogía".

7. Impulsar una discusión sobre el rol de los estándares en la educación, desterrando prejuicios que los identifican con prácticas de enseñanza uniformizadas y rutinarias inconsistentes con concepciones modernas sobre el aprendizaje, con el rol y la imagen profesional de los docentes y con las posibilidades de una educación que dé igualdad de oportunidades a quienes llegan al sistema educativo en distintas condiciones.

8. Considerar la utilidad de formulación de estándares para la selección de candidatos a formación magisterial, para las competencias que deben adquirir durante su preparación inicial y durante su capacitación posterior en servicio, para las condiciones en que se puede ofrecer servicios de formación y capacitación, para la certificación inicial y renovada de competencias profesionales docentes, etc. y el diseño de modelos de evaluación de desempeño docente consistentes o "alineados" con dichos estándares. En dicho proceso deberían tener un rol protagónico organizaciones docentes.

9. Mantener observatorios y promover foros permanentes respecto a tendencias internacionales (y, obviamente, locales) en lo referente al establecimiento de estándares y sistemas de evaluación de aprendizajes, tanto escolares como extra-escolares.

Referencias

Bennett, Randy Elliot (1998). **Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing.**. New Jersey, Educational Testing Service Policy Information Center, Research Division.

Benveniste, Luis (1999). A Comparative Analysis of National Assessment Systems in Southern Cone Countries. Stanford (ensayo de tesis doctoral).

Comisión Centroamericana para la Reforma Educativa del Diálogo Interamericano y el PREAL (2000). **Mañana es muy tarde**.

Comisión Internacional sobre Educación, Equidad y Competitividad Económica del Diálogo Interamericano y el PREAL (1998). **El futuro está en juego**. Washington y Santiago de Chile.

Council of Chief State School Officers (1996). **Trends in State Student Assessment Programs. Data on Statewide Student Assessment Programs.**

CPRE (2000). Consortium for Policy Research in Education / Consorcio de Investigación de Políticas en Educación. "Desarrollando estándares de contenido: creando un proceso para el cambio", próximo a publicarse en la página web del Grupo de Trabajo sobre Estándares y Evaluación de GRADE/PREAL: http://www.grade.org.pe/gtee-preal

Delors, Jacques et al. (1996). **Learning: the Treasure Within, a Report to UNESCO of the International Commission for the Twenty-first Century.** Paris, UNESCO.

Dorn, Sherman (1998). "The Political Legacy of School Accountability Systems" in **Educational Policy Analysis Archives** 6,1.

Ferrer, J. Guillermo (1999). **Aspectos del curriculum intencional en América Latina: Revisión de tendencias contemporáneas en curriculum, indicadores de logro, estándares y otros instrumentos.** Lima, Grupo de Trabajo sobre Estándares y Evaluación de GRADE/PREAL.

Horn, Robin; Laurence Wolff; Eduardo Vélez (1991): **Establecimiento de sistemas de medición del rendimiento académico en América Latina. Un análisis de los problemas y la experiencia más reciente.** Washington, The World Bank, mimeo (RAE-REDUC).

Ravela, Pedro (ed), Wolfe, Valverde, Esquivel (1999). **Los próximos pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?**

Tiana, Alejandro (2000)**. Cooperación Internacional en evaluación de la educación en América Latina y el Caribe: análisis de la situación y propuestas de actuación.** Ministerio de Educación del Brasil/INEP. Documento presentado ante la comisión de seguimiento de los acuerdos de la línea 2 de la Cumbre de las Américas.

UNESCO, PNUD, FNUAP (2000). **Educación para todos en las Américas. Marco de Acción Regional.** Santo Domingo, 10-12 febrero.